



JAPANESE PATENT OFFICE

## PATENT ABSTRACTS OF JAPAN

(11)Publication number: 10187880

(43)Date of publication of application: 21.07.1998

(51)Int.Cl.

G06F 17/21

G06F 12/00

G06F 17/27

G06F 17/30

(21)Application number: 08354966

(71)Applicant:

NEC CORP

(22)Date of filing: 20.12.1996

(72)Inventor:

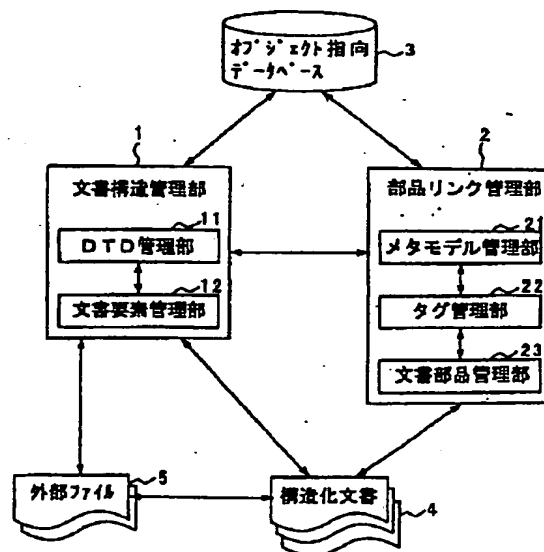
KITANO TAKUYA

(54) DOCUMENT REPOSITORY DEVICE MANAGED BY WORD, SENTENCE AND GRAIN DEGREE OF PART

(57)Abstract:

**PROBLEM TO BE SOLVED:** To retrieve or edit partially necessary information by the unit of a document part concerning a structured document by managing the definition of document type definition(DTD) based on a scheme and managing a document part stuck with a tag defined by DTD based on the scheme.

**SOLUTION:** An object-oriented data base 3 is introduced to a document repository device. Then, in a document structure managing part 1 managing the constitution (chapter number, chapter headline, paragraph, a graphic, etc.) of a structure document 4 based on DTD of standard generalized mark up language(SGML), a DTD managing part 11 manages the definition of DTD based on a prescribed scheme and a document element managing part 12 manages a document part stuck with a tag defined by DTD based on the prescribed scheme. In addition, a part linkage managing part 2 manages the referring relation of a word, a sentence and the grain degree of a part separately from the constitution of the document 4.



③ 2  
(b)

(19) 日本国特許庁 (JP)

# (12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平 10-187680

(43) 公開日 平成10年(1998)7月21日

(51) Int. Cl. °

識別記号

F I

G 0 6 F 17/21

G 0 6 F 15/20 5 7 0 D

12/00 5 4 7

12/00 5 4 7 H

17/27

15/20 5 5 0 E

17/30

15/40 3 4 0

3 7 0 A

審査請求

有

請求項の数 5

F D

(全 14 頁) 最終頁に続く

(21) 出願番号 特願平 8-354966

(22) 出願日 平成8年(1996)12月20日

(71) 出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 北野 拓哉

東京都港区芝五丁目7番1号 日本電気株式会社内

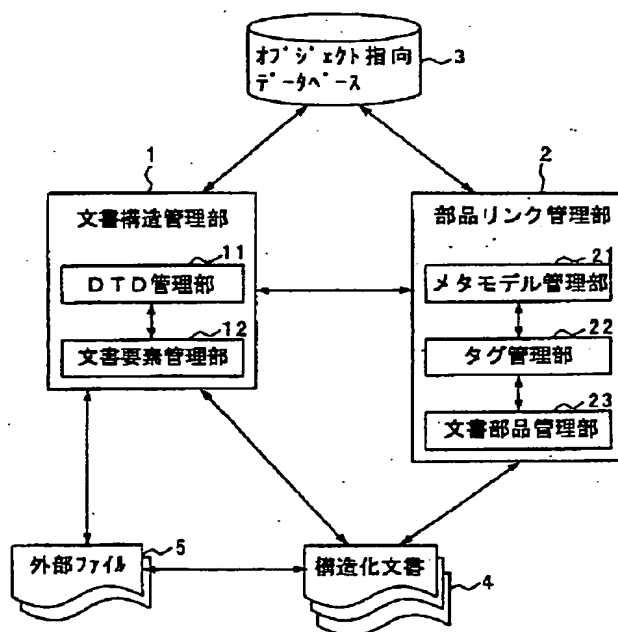
(74) 代理人 弁理士 加藤 朝道

(54) 【発明の名称】 単語、文、部分の粒度で管理するドキュメントリポジトリ装置

(57) 【要約】

【課題】 文書部品間のナビゲート検索、文書部品の削除、追加、更新等の一貫性管理、文書部品の一括編集可能なドキュメントリポジトリ装置の提供。

【解決手段】 SGMLのDTDに基づく構造化文書を管理するためDTDの定義をスキーマを基に管理するDTD管理部11、DTDで定義したタグが張付いた文書部品をスキーマを基に管理する文書要素管理部12、SGMLのDTDを基に構造化文書中の単語、文、部分の粒度で他の構造化文書とリンクを張り管理するためリンクの関係のモデルをスキーマとして管理するメタモデル管理部21、スキーマに従って新にリンクを張るためのタグ定義と管理を行うタグ管理部22、タグを張付けた文書部品とそのリンク関係、構造化文書内の文書部品の出現位置を管理する文書部品管理部23、文書部品のスキーマ定義とこのインスタンスとして管理するオブジェクト指向DB 3を備える。



## 【特許請求の範囲】

【請求項1】 SGML (Standard Generalized Markup Language) の DTD (Document Type Definition) に基づいた構造化文書を管理するために、前記DTDの定義を所定のスキーマに基づいて管理するDTD管理手段と、

前記DTDで定義したタグが張り付いた文書部品を所定のスキーマに基づいて管理する文書要素管理手段と、を含むことを特徴とするドキュメントリポジトリ装置。

【請求項2】 SGMLのDTDに基づいた構造化文書中に含まれる単語、文、部分の粒度で他の構造化文書とリンクを張り管理するために、前記リンクの関係のモデルをスキーマとして管理するメタモデル管理手段と、前記スキーマにしたがって新たにリンクを張るためのタグの定義と管理を行うタグ管理手段と、

前記タグを張り付けた文書部品とそのリンク関係、構造化文書内の前記文書部品の出現位置を管理する文書部品管理手段と、

を含むことを特徴とするドキュメントリポジトリ装置。

【請求項3】 SGML (Standard Generalized Markup Language) の DTD (Document Type Definition) に基づいた構造化文書を管理する装置であって、

前記DTDの定義を所定のスキーマに基づいて管理するDTD管理手段と、

前記DTDで定義したタグが張り付いた文書部品を所定のスキーマに基づいて管理する文書要素管理手段と、を備えた文書構造管理部と、

SGMLのDTDに基づいた構造化文書中に含まれる単語、文、部分の粒度で他の構造化文書とリンクを張り管理するために、前記リンクの関係のモデルをスキーマとして管理するメタモデル管理手段と、

前記スキーマにしたがって新たにリンクを張るためのタグの定義と管理を行うタグ管理手段と、

前記タグを張り付けた文書部品とそのリンク関係、及び構造化文書内の前記文書部品の出現位置を管理する文書部品管理手段と、

を備えた部品リンク管理部と、

を含み、

文書部品のスキーマ定義とこれをインスタンスとして管理を行うオブジェクト指向データベースを用いて、文書部品間のナビゲート検索、文書部品の削除、追加、更新に関する一貫性管理、及び文書部品の一括編集を可能としたことを特徴とするドキュメントリポジトリ装置。

【請求項4】 テキスト形式でない構造化文書外のバイナリ形式の外部ファイルに対してインデックスを付け、前記インデックスの内容と該インデックスが指すバイナリ形式の外部ファイルとの関係を記録し、

バイナリ形式でも請求項1から3のいずれかーに記載の文書部品管理方式と同様の枠組で管理を行うことができるようにしたことを特徴とするドキュメントリポジトリ装置。

【請求項5】 前記DTD管理手段が、タグとタグ属性のスキーマ定義と、そのインスタンスである個々のタグやタグ属性を、オブジェクト指向データベースを用いて管理し、

前記文書部品管理手段が、タグとタグ属性、文書部品と該文書部品がもつタグ属性の値のスキーマを基に、文書部品と文書部品属性のインスタンスを生成して、前記オブジェクト指向データベースを用いて管理することを特徴とする請求項1又は3記載のドキュメントリポジトリ装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、文書中の単語、文、部分にタグを張り付けて文書部品とし、その複数の文書部品がある構造に従って組み立てられた構造化文書を管理するドキュメントリポジトリ装置に関する。

【0002】 後に詳細に説明されるように、本発明のドキュメントリポジトリ (document repository) 装置で行う管理とは、管理対象の構造化文書の文書部品間の構造の定義とその保持であり、この管理により、前記構造に従った文書部品のナビゲート検索、一貫性保持、一括編集を可能とする。

【0003】 本発明のドキュメントリポジトリ装置は、またテキスト形式の文書ファイルとバイナリ形式のファイルを、バイナリファイルにインデックスを付けてそのインデックスをテキストとして扱い、統一的に管理することを可能としたものである。

## 【0004】

【従来の技術】 文書構造を規定する基本規格にODA (Open Document Architecture) がある。これは、文書構造を論理的な視点、割り付け的な視点の2つの視点で、文書全体を木構造の形でまとめるものである。

【0005】 論理的な視点では、段落、図などの構成単位を基本論理対象体とし、これらを集めて章などを構成し、最終的には文書全体となる木構造として文書の論理構造とする。論理構造を持つことによって、段落、章などの論理単位での編集や交換時の論理単位での保護、論理単位に対する自動番号付けなどの支援が可能となる。

【0006】 一方、割り付け的な視点では、ページ、頭書、枠 (図や表のこと)、区画などの単位を集め、論理的な視点と同様に、文書全体を木構造として表現する。

【0007】 また文書構造に関する他の規格としてSGML (Standard Generalized Markup Language; スタンダードジェネラライズドマークアップ言語) がある。これは文書型定

義DTD (Document Type Definition; ドキュメントタイプデフィニション) で、テキストの文書要素 (見出し、本文、リストなど) に張り付けるタグを定義する。DTDでは、このタグの親子関係、出現順序、出現回数などを木構造で定義することができ、この文法に従ったSGMLで書かれるテキストは、構造化文書と呼ばれる。

【0008】図2に、手紙を例にとったDTDの木構造表記の一例を示す。図2を参照すると、タグ「letter」で囲まれた文書部品は、タグ「date」(日付)、「to」(宛名)、「body」(本文)、「from」(差出人)、「ps」(追伸)、をそれぞれ持つ子部品を持つことを示している。

【0009】これら子部品は、前記した順序通りに出現するものとし、必ず1回出現しなければならない。最後の「ps」タグの文書部品には「?」記号が付いており、0回または1回出現することを表している。

【0010】また「body」タグで囲まれる文書部品は、「p」(段落)タグを持つ子部品を持ち、出現回数は「+」記号で1回以上を表す。

【0011】「p」タグで囲まれる文書部品は、「hp」(強調)タグを持つ子部品を持ち、出現回数は「\*」記号で0回以上を表す。

【0012】ODAもSGMLも構成要素の木構造で文書表現する。異なる点は、ODAは標準言語の一つであるのに対し、SGMLはDTDのタグの定義による。文書を記述する言語のためのメタ言語である点である。

【0013】2つの標準をベースにした構造化文書処理装置に関連する従来技術の一例として、例えば特開平6-301721号公報には、論理的構造を持った文書の全文検索のための、リレーショナルデータベース(RDB)を用いたSGML文書の管理が記載されている。

【0014】上記特開平6-301721号公報が提案する方法では、SGMLで記載された原文書を入力データとし、そのSGMLに基づいて、その構造情報、タグ情報、テキスト情報の各テーブルをDDL (Data Definition Language) を用いて定義する。そしてそれぞれのテーブルを生成し、RDBに格納する。SGMLを用いた原文書データのデータも格納する。文書の検索は、検索言語DQL (Document Query Language) を用いて文書構造を指定した検索要求の入力を受け付け、リレーショナルデータベース管理システム(RDBMS)に対してSQL書式の検索要求を行う。検索結果の表示は、前記検索要求の結果を、RDB中の前記構造情報を用いてタグを含むSGMLの形式に編成して表示する。

【0015】また例えば特開平6-314281号公報には、異なる文書処理システム間で文書データを交換する際などに利用される構造化文書について、その特定論理構造からデータ量の少ない共通論理構造を持った構造

化文書を生成する装置が提案されており、ODAなどで述べられている共通論理構造の生成にあたって、オブジェクト指向の考えを導入している。

【0016】すなわち上記特開平6-314281号公報に提案される文書編集装置においては、ある特定の論理構造を基に、その各オブジェクトと1対1に対応するクラスを生成し、各クラスの属性は対応するオブジェクトからコピーし、各クラス間の階層構造は対応するオブジェクト間の階層構造をそのまま対応付けて共通論理構造のクラスを生成する文書編集装置について述べている。クラスは例えば章番号、章見出し、段落、図形などの各文書部品に対応して与えられる。また前記4つの部品を下位クラスとし、これを集約させ上位クラスの章クラスを与えるなど、クラスの部分-集約関連を導入している。

【0017】そして例えば特開平6-119331号公報には、構造化文書に対して文書内の枠や見出しによる検索を可能にする文書部品検索装置が提案されている。これは、文書部品の論理構造を特定し、その特定した論理構造にインデックスを付けて管理する。枠や見出しによる検索は、その枠や見出しに相当する文書部品と、前記付加したインデックスを検索して得た文書部品の属性に関する情報を条件照合して行う。具体例としては、3つの文書A、B、Cの文書の内部にそれぞれ枠を持ち、インデックスとして表枠、図形枠、グラフ枠を保持する記憶装置に対する検索が示されている。検索としては、文書部品「枠」を文書内部に持ち、それがグラフ枠である文書の検索を行っている。

【0018】ODAやSGMLなどの文書構造化の技術の他に、本発明に関連する技術として、ハイパーテキストがある。これはWWW (World Wide Web; ワールドワイドウェブ) ブラウザ上へのリソース表示に用いられるHTML (Hyper Text Markup Language; ハイパーテキストマークアップ言語) が代表的であり、HTMLはSGMLのある1つのDTDによって実現されている。

【0019】HTMLは、WWWブラウザ上にアンカーと呼ばれるマウスでクリックすることのできる部分を作ることができる。このアンカーに、別のHTMLリソースや同じHTMLリソースの別の場所を指し示すようなリンクを指定し、アンカーのマウスクリックによって、その場所に移動させて表示させることができる。

【0020】このリンクの指定は、インターネット上のあらゆるサーバマシンのリソース (ただし、セキュリティによってアクセス制限されているサーバマシンやリソースは除く) に対して可能であり、これにより、現在のWWWの世界を築いている。

【0021】

【発明が解決しようとする課題】しかしながら、上記した従来技術は、下記記載の問題点を有している。

【0022】(1) まず上記特開平6-301721号公報に提案される方法においては、データベースにRDB(リレーショナルデータベース)を導入している。構造化文書中の文書部品の構造は、親子関係、包含関係を表す木構造や、参照関係を表すネットワーク構造である。しかしながら、RDBが扱うテーブル形式では、このような構造を表現するには、ギャップが大きい。

【0023】またテーブル間にまたがる文書部品の検索要求には、RDBではテーブル間の結合演算を行う必要があり、一般にこの演算には多くの計算時間を要する。

【0024】(2) 次に上記特開平6-314281号公報に提案される文書編集装置では、部品の関係を部分、集約の関係のクラス化のみにとどまっており、クラス間の一般的な関係、参照関係に関するクラスのモデル化については、記載されていない。複数の構造化文書にまたがった、例えばある文書部品は別の文書部品と関連があり、検索でその関連をたどれるようにリンク付けするような共通論理構造の生成は考えられていない。このため、構造化文書間をリンクでたどるナビゲート検索を可能にする構造化文書のオブジェクトを生成することはできない。

【0025】(3) さらに上記特開平6-119331号公報に提案される文書部品検索装置は、枠に付けられている属性が例えば幾何学図形、ラスタ図形、表、数式などで、その属性を指定して前記属性を含む文書の検索を行うものである。

【0026】しかしながら、枠の内部にまで検索範囲は及ばず、例えば表の列の見出しが何であるかを検索キーとして、その見出し(検索キー)を含む表をまた含む文書の検索は不可能である。

【0027】これは、上記上記特開平6-119331号公報の文書部品検索装置においては、インデックスを枠全体に対して1つだけ付けていて、しかもそのインデックスも枠の種類に注目したものであることが、前記した検索能力の限界につながっている。

【0028】(4) ところで、WWWに代表されるHTMLのハイパーテキスト技術は、そのリンクの指定を自由に設定できることが大きな利点であるが、その反面、構造的でない。

【0029】ここで、「構造的でない」とは、あるリソースと他のリソースとが、ある関係を持つという定義を前提にして、リンクが設定されていない、ということである。

【0030】自由にリンクを設定できる環境において、リンク関係の管理を集中的に行おうとした場合、極めて困難なものとなる。すなわち、ある時点であるリソースからのリンク先をすべて把握しても、そのリンク先のリソースが消えてしまったり、新たなリソースへのリンクが追加されてしまったり、という事態も予想され、結局、管仕切れない、ということである。

【0031】したがって、本発明は、上記問題点に鑑みてなされたものであって、その目的は、構造化文書をある規定した構造で管理し、管理した構造化文書を文書部品単位で部分的に必要な情報を検索したり、あるいは編集したりすることを可能とする、ドキュメントリポジトリ装置を提供することにある。

【0032】また、本発明の他の目的は、バイナリファイルなど、その中身にタグを張り付けることができない文書に対してインデックスを与えて管理することを可能とするドキュメントリポジトリ装置を提供することにある。

【0033】

【課題を解決するための手段】前記目的を達成するため、本願第1発明は、SGML(Standard Generalized Markup Language)のDTD(Document Type Definition)に基づいた構造化文書を管理するために、前記DTDの定義をスキーマに基づいて管理するDTD管理手段と、前記DTDで定義したタグが張り付いた文書部品をスキーマに基づいて管理する文書要素管理手段と、を含むことを特徴とする。

【0034】また、本願第2発明は、SGMLのDTDに基づいた構造化文書中に含まれる単語、文、部分の粒度で他の構造化文書とリンクを張り管理するために、前記リンクの関係のモデルをスキーマとして管理するメタモデル管理手段と、前記スキーマにしたがって新たにリンクを張るためのタグの定義と管理を行うタグ管理手段と、前記タグを張り付けた文書部品とそのリンク関係、構造化文書内の前記文書部品の出現位置を管理する文書部品管理手段と、を含むことを特徴とする。

【0035】また、本願第3発明は、本願第1発明における前記DTD管理手段、及び前記文書要素管理手段と、本願第2発明における前記メタモデル管理手段、前記タグ管理手段、前記文書部品管理手段の、文書部品のスキーマ定義と、これをインスタンスとして管理を行うオブジェクト指向データベースを用いて、文書部品間のナビゲート検索、文書部品の削除、追加、更新に関する一貫性管理、及び文書部品の一括編集を可能としたことを特徴とする。

【0036】そして本願第4発明は、テキスト形式でない構造化文書外のバイナリ形式の外部ファイルに対してインデックスを付け、前記インデックスの内容と該インデックスが指すバイナリ形式の外部ファイルとの関係を記録し、バイナリ形式でも本願第1から第3のいずれかに記載の文書部品管理方式と同様の枠組で管理を行うことができるようにしたことを特徴とする。

【0037】

【発明の実施の形態】本発明の実施の形態について図面を参照して以下に説明する。図1は、本発明の実施の形態のドキュメントリポジトリ装置の基本構成を示すプロ

ック図である。

【0038】図1を参照すると、本発明は、その好ましい実施の形態において、オブジェクト指向データベース (Object Oriented Data Base ; 「OODB」という) 3をドキュメントリポジトリ装置に導入する。このOODBは、文書部品のつながりを表す概念モデルとオブジェクト指向のスキーマ実装のセマンティクスギャップが少ない。

【0039】また本発明の実施の形態においては、OODBを用いてオブジェクトのつながりを予めデータベースに記憶する方式をとるので、RDB (リレーショナルデータベース) のようにSQLの発行によって検索実行時にテーブル間のつながりを計算する方式よりも、オブジェクト間のナビゲーション検索が高速になる。これにより、上記特開平6-301721号公報に関する上記した問題点(1)を解消している。

【0040】また上記特開平6-314281号公報に関する上記問題点(2)とWWWに代表されるHTMLのハイパーテキスト技術の上記問題点(4)を解消するために、本発明は、その好ましい実施の形態において、図1を参照すると、SGMLのDTDに基づいた文書の構成(章番号、章見出し、段落、図形など)を管理する文書構造管理部1と、前記文書の構成とは別に、文書中の単語、文、部分の粒度の参照関係を管理する部品リンク管理部2を備える。

【0041】文書構造管理部1は、DTD管理部11と文書要素管理部12から構成される。

【0042】また部品リンク管理部2は、メタモデル管理部21と、タグ管理部22と、文書部品管理部23と、から構成される。

【0043】構造化文書4は、これら5つの管理装置を通して、OODBに登録される。

【0044】図3は、本発明の実施の形態の動作を説明するための図であり、この構造化文書OODBへの登録における、5つの管理装置(すなわちDTB管理部11、文書要素管理部12、メタモデル管理部21、タグ管理部22、文書部品管理部23)の全体の流れと個々の働きを示している。

【0045】図3を参照すると、文書構造管理部1も部品リンク管理部2もそれぞれDTDの決定111やメタモデルの定義211というような、共通構造のクラス生成をまず第一に行う。

【0046】このアプローチは、従来のハイパーテキストと異なるアプローチであり、文書内のリンク構造を予めメタモデルで定義し、そのモデルをOODBのスキーマとして管理する。すなわち、そのスキーマの範囲を越えたリソース間のリンクを許さないわけである。

【0047】このことは、結果として、ハイパーリンクの自由度を下げることになるが、前記したように、ハイパーテキストでは困難であったリソースのリンク関係の

集中管理を本発明のドキュメントリポジトリ装置は可能としている。

【0048】また予め定義されたメタモデルの構造を活かしたナビゲーションも、本発明のドキュメントリポジトリ装置では可能になる。

【0049】上記したアプローチは、上記特開平6-314281号公報に記載される、既存の構造化文書一つ一つから共通論理構造のクラスを生成するアプローチとも異なる。生成しようとする構造化文書の共通論理構造を最初に生成し、これを用いて各構造化文書のオブジェクトを生成し、管理するアプローチが、本発明のアプローチであるからである。

【0050】構造化文書をOODBに登録するためには、図3を参照すると、まずDTDの決定処理111で、SGMLのDTD定義を用意する。そして、タグとその木構造のスキーマ定義処理112で、タグをクラスとみなしたオブジェクトモデルで表現し、これを基に、OODBのスキーマ生成を行う。

【0051】図4は、そのスキーマを示したものあり、DTD定義に含まれるタグとタグの属性、そしてタグの親子関係をモデル化している。なお図4は、OMT (Object Modeling Technique ; オブジェクトモデリングテクニック) の表記法に従った。

【0052】DTD管理部11は、図4に示す、タグ31とタグ属性32のスキーマ定義(クラス間の関連を含む)と、そのインスタンスである個々のタグやタグ属性をOODBを用いて管理する。

【0053】次に、文書インスタンスの作成処理121で、DTDの決定処理111で決定したDTDに従った文書インスタンスを作成する。文書部品のDB登録122では、前記文書インスタンス中の文書部品を、タグとその木構造のスキーマ定義処理112で生成したタグとタグ属性のスキーマとの関係付けを行う。

【0054】その際、DTDの決定処理111で定義したタグとその属性を、属性値を決めて文書部品に張り付けていくというイメージの処理が行われる。

【0055】タグとタグ属性、文書部品とその文書部品もつタグ属性の値の関係(スキーマ)の一例を、図5に示す。

【0056】文書要素管理部12(図1参照)は、図5に示すスキーマを基に、文書部品33と文書部品属性34のインスタンスを生成して、OODBで管理する。

【0057】このインスタンス管理により、実際の文書インスタンス中の前記スキーマのインスタンスに対応する文書部品の管理が行われる。

【0058】以上のDTDの決定処理111、タグと木構造のスキーマ定義1処理12、文書インスタンスの作成処理121、文書部品のDB登録処理122のステップで格納された構造化文書を、その構造化文書中の単

語、文、部分の粒度でリンクを張って管理する場合、以下で説明する部品リンク管理部3を用いて管理する。

【0059】まず、メタモデルの定義211で、構造化文書全体のリンク関係のモデルを定義する。このリンク関係のモデルの定義には、OMTなどに代表されるオブジェクト指向分析の手法を用いることができる。すなわち、タグを一つのクラスに対応付け、そのクラスを用いてタグ同士の一般的な参照関連、部分、集約関係、汎化-特化の関連、関連の多重度などを定義することができる。

【0060】またクラスに属性や操作を加えて、そのクラスに属する文書部品に共通な操作を記述することもできる。例えば、同じタグの付いた（すなわち同じクラスに属した）文書部品に対して一括編集を行う操作を記述したり、特殊な属性を加えて操作に対する制約を与えたり、さらにある文書部品からのナビゲート検索のための操作を記述したりとさまざまな定義をメタモデルに与えることができる。このタグのリンク関係のモデルのためのクラスを抽象的に「タグリンク」と呼ぶことにすると、リンクタグは、図5で示したスキーマのタグ31を継承（inherit）し、図6に示すスキーマの構成要素クラスとなる。

【0061】図6を参照して、タグリンク35の属性であるidは、タグ属性32の一つで、名前がid、型が正数である。idは、リンクタグ35のインスタンス（すなわち実際のタグ）に通し番号の属性を付けて識別するためにある。

【0062】再び図3を参照して、メタモデルのスキーマ定義212では、メタモデル定義211で定義したメタモデル（オブジェクトモデル）をOODBのスキーマへの変換を行う。

【0063】メタモデル管理部21は前記変換を行ったスキーマのスキーマ生成を行い、OODBに登録する。

【0064】メタモデルを基にしたタグの定義221では、メタモデルで定義した各クラスに対して、実際に文書インスタンスに張り付けるタグの名前（文字列）を定義する。

【0065】タグのDB登録222では、タグ管理部22が、図6に示すメタモデルのリンクタグに相当する各クラスのインスタンスを生成する。このインスタンス生成の引数として、メタモデルを基にしたタグの定義221で定義したタグの名前（文字列）を与える。これにより名前が決定した各タグがOODBに登録される。

【0066】文書に対するタグの張り付け231では、メタモデルを基にしたタグの定義221で定義したタグを実際の文書インスタンス中に張り付け、単語、文、部分の粒度の文書部品を作っていく。文書部品のDB登録232では、図7に示す文書部品33を継承して付け加えたリンク部品36のインスタンスを生成していく。

【0067】図7を参照して、リンク部品36のインス

タンスは、タグを定義したリンクタグ35のインスタンスへのポインタ（関係37のインスタンス）を持ち、このポインタが、リンク部品36のインスタンス（すなわち文書インスタンス中の文書部品）にどのタグが張り付いているかを識別する。

【0068】またリンク部品36のインスタンスは、関係38が表すように、自分がどの文書部品を親に持つかを構造化文書全体の構造木に従い記録する。

【0069】文書部品管理部23は、上記関係37、38のリンクを含んだリンク部品36のインスタンス生成とOODBへの登録を行い、単語、文、部分の粒度でリンク関係を張られた文書部品を管理する。

【0070】WWWに代表されるHTMLのハイパーテキスト技術における、リンク関係の管理に関する問題を解決するために、以上説明した通り、文書部品のリンク関係も含めたインスタンス生成を行いOODBに登録している。

【0071】ある文書部品を削除したりあるいは追加したりする場合、その文書部分は基になるメタモデルに従わなければならない。

【0072】削除に関しては存在制約がある。リンクに関して参照先と参照元のそれぞれの文書部品があり、それがインスタンスとしてOODBに登録されている場合、参照先のインスタンスの削除は、参照元のインスタンスの属性を更新することになる。OODBは、あるインスタンスの削除に関する他のインスタンスへの影響を感知することができる（この機能が実装されている）。このため、OODBを用いた本発明のドキュメントリポジトリ装置は、データベース操作を行うユーザに対して注意を促したり、削除を禁止させたりすることができる。

【0073】そして文書部品の追加に関しては、メタモデルが示す関連の有無やカーディナリティ制約などがある。これは、メタモデルで定義されていないクラス間でそのインスタンス間のリンクを張ることはできない、あるクラスに関連する他のクラスが生成できるインスタンスの数を越えてインスタンスを生成することはできないなどである。

【0074】ある文書部品をどうしてもドキュメントリポジトリ装置に格納したい場合には、その文書部品の追加に対応するようメタモデルを修正し、OODBのスキーマ拡張を行う必要がある。

【0075】次に上記特開平6-119331号公報に記載の従来技術の上記問題点（3）を解消するために、本発明の実施の形態においては、枠に対するインデックスは文書インスタンス中の文書部品とし、この文書部品から枠に対するリンクを必要な数だけ張り、枠の持つ内部の情報をそのリンク元である文書部品の内容で表す方法を探る。ワープロなどの文書編集エディタによって、図や表やグラフをテキスト形式で表現するものもあ

るが、一般には、図や表やグラフは別の専用エディタを用いて編集し、バイナリ形式の外部ファイルとして文書の中からこれらを参照する形になっている。

【0076】本発明の実施の形態では、外部ファイルの参照部分に張り付けるタグとそのインデックス部分を表すタグとの関係をメタモデルで表現する。図8は、このメタモデルの一例を示す図であり、図中、41、42、43の部分がメタモデルである。なおメタモデルは、図7では、リンクタグ35に相当する。

【0077】図8を参照して、それぞれのメタモデルのクラス(41と42)からインスタンスを生成し、タグの名前を決定する(<IDX id="1">441、<IDX id="2">442、<GAIBU id="1">45)。

【0078】なお、このインスタンス生成時に、メタモデルのクラス間の関係43により、各インスタンス間のリンクも張られる。このようにして生成したタグを、文書インスタンス中のインデックス部分と外部ファイル参照部分に張り付け、新たな文書部品を作成する。図9に、新たな文書部品の一例を示す。

【0079】図9を参照して、図7に示すリンク部品36のインスタンスである文書部品「部屋の間取り」51が、タグ441(<IDX id="1">)によってタグ付けされ、同じく文書部品「2DK」52、「部屋2つの台所1つ」53がタグ442(<IDX id="2">)によってタグ付けされている。

【0080】同様に、文書部品「<IMGSRC="madori.gif">」54は、タグ45(<GAIBU id="1">)によってタグ付けされている。

【0081】なお、図9において、「2DK」52と「部屋2つ台所1つ」53に同じタグ(<IDX id="2">)が付いている意味は、リンク部品36の属性である内容が違っても、リンクタグ35のレベルでは同じ扱いをするという意味である。同義語の扱いを同じにするなどが例としてある。

【0082】リンクタグ35とリンク部品36のそれぞれのインスタンスは、部品リンク管理(図1の2)によって管理され、これによりバイナリファイルなどの外部ファイルへのインデックス付けや検索などが可能になる。

【0083】このように本発明の実施の形態に係るドキュメントリポジトリ装置は、構造化文書のある規定した構造で管理し、管理した構造化文書を文書部品単位で部分的に必要な情報を検索したり、あるいは編集したりする場合に用いて好適とされる。またバイナリファイルなど、その中身にタグを張り付けることができない文書に対して、インデックスを与えて管理する場合に用いて好適とされる。

【0084】

【実施例】上記した本発明の実施の形態について更に詳

細に説明すべく、本発明の実施例について以下に説明する。

【0085】[文書データベース登録]本発明の一実施例として文書データベース(OODB)への登録について説明する。図10に、本発明の一実施例として、DTD木構造に基づいた文書の登録の例を示す。ここでタグの属性は特にないものとする。このDTD定義のオブジェクト指向データベース(OODB)への登録には、図4に示したスキーマモデルを用いる。

10 【0086】図10を参照して、タグ61~68はそれぞれ名前が、「doc(文書)」、「sec(章)」、「subsec(節)」、「p(段落)」、「hp(強調)」、「fig(図)」、「list(箇条書)」、「item(箇条)」であり、図5に示すタグ31のインスタンスとして、DTD管理部11(図1参照)によって、OODB3に登録される。

【0087】またタグの親子関係も、図10に示す通りの関係で、OODB3(図1参照)に登録される。

20 【0088】次に前記定義したタグを実際の文書中に張り付け、文書部品のインスタンスを生成してOODB3に登録する。図11は、その文書インスタンスの例を示す図である。

【0089】文書にタグを張り付けるとき、文書要素管理部はそのタグ付けした文書部品を図5に示すスキーマモデルに従って、文書部品33のインスタンスとしてOODB3に登録する。また文書部品レベルの親子関係38もOODB3(図1参照)に登録する。

30 【0090】[構造化文書の部品リンクの登録]次に、本発明の一実施例として、構造化文書の部品リンクの登録について説明する。図11に示す構造化文書を例に、単語、文、部分の粒度の文書部品間のリンク関係の登録の手順を示す。

【0091】図11は、エレクトリックコマースの注文処理を実現させるシステムの基本仕様書の一部である。コンピュータ端末上に表示されるGUI(グラフィカルユーザインタフェース)を通してユーザから注文を受け、システムはその注文の伝票作成と、商品の在庫更新を行う。

40 【0092】基本仕様書にはオブジェクト指向分析/設計で用いられるクラス、関連、属性、操作を表す単語を含んでおり、これらの単語はある構造的な関係を持っているとする。

【0093】また図11とは別に、バイナリ形式の外部ファイルとして注文処理を行うGUIのイメージ図と、クラスや属性などの関係を表すオブジェクト図があると、これらバイナリ形式の図と、図11に示す構造化文書中の単語の関係を持たせ、全体で構造化関係を持たせることにする。

50 【0094】まず、メタモデルの定義を行う。メタモデルの定義に用いるスキーマクラス(タブの定義のために



モデル化したクラス)は、図6に示したリンクタグ35に相当し、これから定義するメタモデルのクラスはすべてタグ31を継承する。

【0095】図12は、エレクトリックコマースの注文処理に関する構造化文書間のメタモデルを示した図である。図12を参照して、71~74は、オブジェクト指向分析/設計で用いられる「クラス」、「関連」、「属性」、「操作」を表し、これらの構造的な関係がモデル化されている。また75~78は、バイナリファイルに関するモデルであり、汎化-特化の関係を用いてモデル化されている。

【0096】図12のクラス71、関連72、属性73、操作74と、「図」76の関係は図8で示したインデックス41と外部ファイル42の間の関係43と同じで、71~74が、「図」76のインデックスになっている。

【0097】図12に示したメタモデルをOODBのスキーマへ変換する。メタモデル管理部は前記変換を行ったスキーマのスキーマ生成を行い、OODBに登録する。

【0098】次に、メタモデルの各クラスに対してタグの名前を与え、インスタンスを生成してOODBに登録する。そのタグの名前を71から順に「cl」、「rl」、「at」、「mt」、「bf」、「fig」、「gui」、「ob」とする。

【0099】作成した前記タグを図11に張り付け、図7のリンク部品36のインスタンスを生成する。

【0100】図13に、図11の文書インスタンスにタグを張り付けた例の一部を示す。図13に現れているタグ付けされた文書部品を整理すると、図14に示すようなものとなる。ただし、図14は、説明の便宜上整理して示した表であり、本発明のドキュメントリポジトリ装置において、このような表形式でOODBに格納しているわけではない。

【0101】図14を参照して、タグにあるidは、図6に示すリンクタグ35の属性である。この例で説明すると、文書部品「商品名」、「価格」、「説明」は、いずれもメタモデルの同じクラス「属性」73(図12参照)のインスタンスであるが、これらがインスタンスレベルで識別されることを、idの通し番号で表現している。

【0102】また文書部品「商品名」と「情報」は同じタグ<at id="1">が付いている。これは、インスタンスレベルで同じ意味であることを示す時に用いる。例えば同義語や、ある意味において文書部品のクラスリングを行う時などに有効である。この関係が、図7における関係37の、一つのリンクタグに対して、複数のリンク部品が存在するというスキーマの実現例である。

【0103】次に、これらの文書部品を、図7に示した

スキーマや、図12に示したメタモデルのインスタンスとして、OODBに登録する。メタモデルを基に、例として図15に示すようなインスタンスの関係を持つものとした。

【0104】文書部品の親子関係(図7の関係38)については、図13の文書インスタンスの例により、前記文書部品すべてが節<subsec>システムのシナリオの第一段落<p>注文フォームには…に含まれている。よって、前記文書部品すべてはこの段落<p>注文フォームには…を親部品に持つ。この関係をOODBに登録し、前記文書部品の文書インスタンス中の存在位置を記録する。

【0105】図12の「図」76とのリンク関係は、図9を参照して説明した方法でリンク関係を設定する。

【0106】図15中のインスタンスのうち、バイナリファイルであるGUIイメージ図やオブジェクト図のインデックスとしたいタグのみを、前記タグの名前を付けた「fig」、「gui」のタグとリンクを張り、OODBに登録する。

【0107】[文書部品の構造に基づいたナビゲーション検索]次に本発明の一実施例として、文書部品の構造に基づいたナビゲーション検索について説明する。図12を例に説明すると、例えば以下に列挙するメタモデルの構造に従った文書部品の検索が可能である。

【0108】・ある「クラス」のインスタンスの持つすべての「属性」、「関連」のインスタンスを挙げよ。

【0109】・ある「クラス」のインスタンスとある「クラス」の「関連」のインスタンスを挙げよ。

【0110】・ある「オブジェクト図」のインスタンスに含まれる(インデックス付けされている)「属性」のインスタンスを持つ「クラス」のインスタンスが持つ「操作」インスタンスを挙げよ。

【0111】図15を例として説明すると、クラス「商品」の属性のインスタンスをすべて挙げよであれば、属性「商品名」、「価格」、「説明」が返ってくる。また図14で、タグ<at id="1">を持つ文書部品インスタンスをすべて挙げよとすれば、文書部品「商品名」、「情報」が返ってくる。

【0112】[文書部品の構造に基づいた一貫性管理]次に本発明の一実施例として、文書部品の構造に基づいた一貫性管理について説明する。図12に示す例に説明すると、存在制約から、ある「クラス」のインスタンスを削除すると、それと関連がある「属性」と「操作」のインスタンスに影響が出てくる。その前記「クラス」のあるインスタンスを削除する場合、そのクラスのインスタンスを参照する前記「属性」と「操作」のインスタンスのリンクを削除しなければならない。また場合によっては、前記「属性」と「操作」のインスタンスそのものを削除する必要もある。

【0113】図15に示す例で説明すると、クラス

「客」を削除すると、その操作である「購入のチェックボックスに…」はもはや必要ない操作であるので、この操作のインスタンスも同時に削除される。

【0114】インスタンスの追加に関する制約では、図12を例に説明すると、「図」76にインデックスとして付けることのできるタグは<c1>、<r1>、<a1>、<m1>のみであり、他の、例えば<b1>などのタグは、「図」76のインデックスとしてリンクを張ることはできない。

【0115】インスタンスの更新に関する制約では、あるインスタンスの更新がそれと関連するインスタンスにとって妥当な関連であることが保証されるかどうかということである。

【0116】図12を例に説明すると、「GUIイメージ図」77のあるインスタンス、すなわちそのGUIイメージ図そのものを大幅に書き換えたとした時、そのインデックスとして付けられていたクラス71、関連72、属性73、操作74のそれぞれのインスタンスがインデックスとして妥当かどうか改めて検討する必要がある。この場合、この「図」76と、71～74の関連には、更新制約が存在することを記録しておき、更新操作が起こった場合、その更新に影響のあるすべてのインスタンスを列挙して一つ一つ更新の確認を行う。

【0117】〔文書部品の構造に基づいた文書部品の一括編集〕次に本発明の一実施例として、文書部品の構造に基づいた文書部品の一括編集について説明する。本発明のドキュメントリポジトリ装置は、メタモデル、タグ、文書部品と、3層の抽象階層を設けて、文書部品を管理している。構造化文書内の文書部品をある規則に従って一括編集する場合、前記3層のどの層からも編集操作が可能である。

【0118】メタモデルのレベルでは、そのメタモデルの構造変更や、あるスキーマクラスに属するすべてのインスタンス（文書部品）の文書部品属性の値を変更するとか、インスタンスすべてを全消去するとかが可能である。

【0119】タグレベル、文書部品レベルもそのレベルで集合する文書部品に対し同様の操作を施すことができる。かつ、これらの操作を、上記一貫性管理に基づいて行うことができる。

【0120】図14に示す例を参照して、一括編集の例を説明すると、文書部品「商品名」と同じタグ、同じid番号のタグの文書部品を「商品名」に書き換えようとすると、構造化文書中の<a1id="1">の付いた文書部品「情報」がすべて「商品名」に置き換わる。

【0121】

【発明の効果】以上説明したように、本発明によれば下記記載の効果を奏する。

【0122】本願第1発明（請求項1）によれば、SGMLのDTDに基づいた構造化文書がOODBによって

管理することができる。

【0123】このように、OODBに管理することによって、構造化文書中の文書部品の親子関係、包含関係を表す木構造、参照関係を表すネットワーク構造などと、人間の直観的イメージとのセマンティクスギャップを埋める。またナビゲート検索に関し、RDBではテーブル間の結合演算方式より高速に検索を行うことができる。

【0124】本願第2発明（請求項2）によれば、構造化文書中の文書部品の親子関係、包含関係という関係の定義に主眼がおかれるSGMLのDTD定義に対し、ハイパーテキスト技術に代表される文書間のリンク関係を、単語、文、部分の粒度で管理することを可能とする。更に、ハイパーテキストのリンクの仕組みに構造的な管理要素を加え、リソースの消滅などによるリンク関係の保持の難しさを解決している。

【0125】本願第3発明（請求項3）によれば、ドキュメントリポジトリ装置が有する構造化文書に対する文書部品間のナビゲート検索、文書部品の削除、追加、更新に関する一貫性管理、文書部品の一括編集の機能により、構造化文書のある定められた規則にしたがった編集や必要情報の抽出などを容易化する。

【0126】本願第4発明（請求項4）によれば、バイナリ形式のファイルを、その中身を表現するインデックスを与えてテキスト形式のファイルと同様に扱う枠組を提供し、構造化文書の管理に統一性を与えるものである。

【図面の簡単な説明】

【図1】本発明の実施の形態に係るドキュメントリポジトリ装置の構成を示す図である。

【図2】手紙のDTD木構造表記を示す図である。

【図3】本発明の実施の形態の処理動作を説明するための図であり、文書構造管理部と文書リンク管理部による文書のデータベース登録の流れを示す図である。

【図4】本発明の実施の形態を説明するための図であり、DTDのタグとタグ属性、タグの親子関係を表すスキーマモデルを示す図である。

【図5】本発明の実施の形態を説明するための図であり、タグ、タグ属性、文書部品、文書部品属性の関係を表すスキーマモデルを示す図である。

【図6】本発明の実施の形態を説明するための図であり、文書構造管理部のスキーマ（フレームワーク）から継承してタグリンクを加えた、メタモデル管理部のスキーマを示す図である。

【図7】本発明の実施の形態を説明するための図であり、文書構造管理部のスキーマ（フレームワーク）から継承してリンク部品を加えた、文書部品管理部のスキーマを示す図である。

【図8】本発明の実施の形態を説明するための図であり、外部ファイルとインデックスの関係を表すメタモデルの一例を示す図である。

【図9】本発明の実施の形態を説明するための図であり、文書インスタンスに対するタグの張り付けの一例を示す図である。

【図10】本発明の実施例を説明するための図であり、DTD木構造の例を示す図である。

【図11】本発明の実施例を説明するための図であり、文書インスタンスの例を示す図である。

【図12】本発明の実施例を説明するための図であり、エレクトリックコマースの注文処理に関する構造化文書間のメタモデルを示す図である。

【図13】本発明の実施例を説明するための図であり、図11の文書インスタンスにタグを張り付けた例の一部を示す図である。

【図14】本発明の実施例を説明するための図であり、タグ付けした文書部品の一覧表を示す図である。

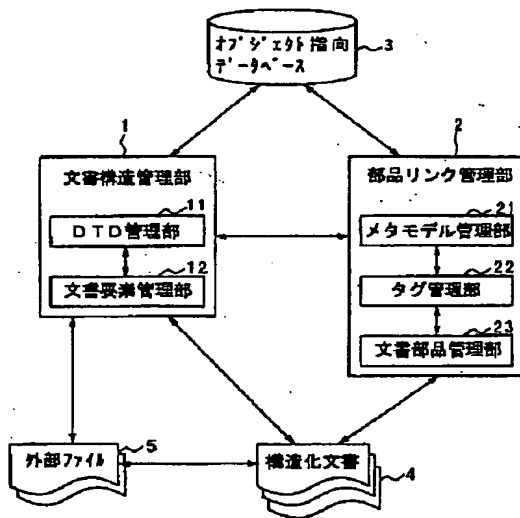
【図15】本発明の実施例を説明するための図であり、メタモデルに従ったタグのインスタンスレベルのリンク

関係の例を示す図である。

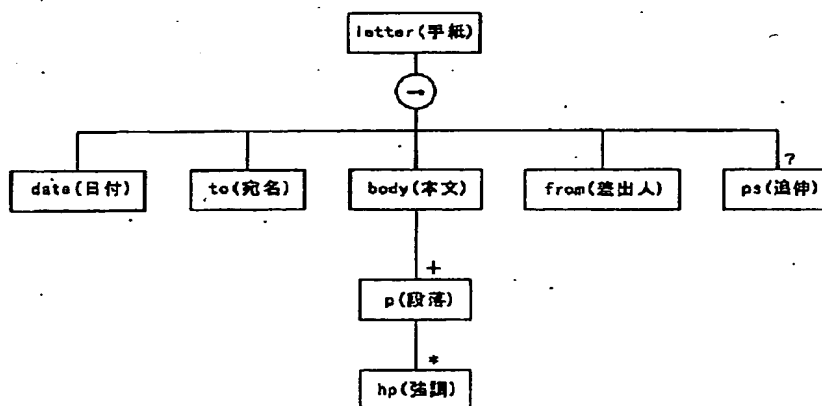
【符号の説明】

- 1 文書構造管理部
- 2 部品リンク管理部
- 3 オブジェクト指向データベース
- 11 DTD管理部
- 12 文書要素管理部
- 21 メタモデル管理部
- 22 タグ管理部
- 23 文書部品管理部
- 31 タグ
- 32 タグ属性
- 111 DTDの決定処理
- 112 スキーマ定義処理
- 121 文書インスタンスの作成処理
- 122 DB登録
- 211 メタモデルの定義

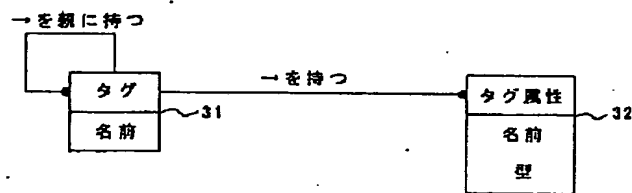
【図1】



【図2】



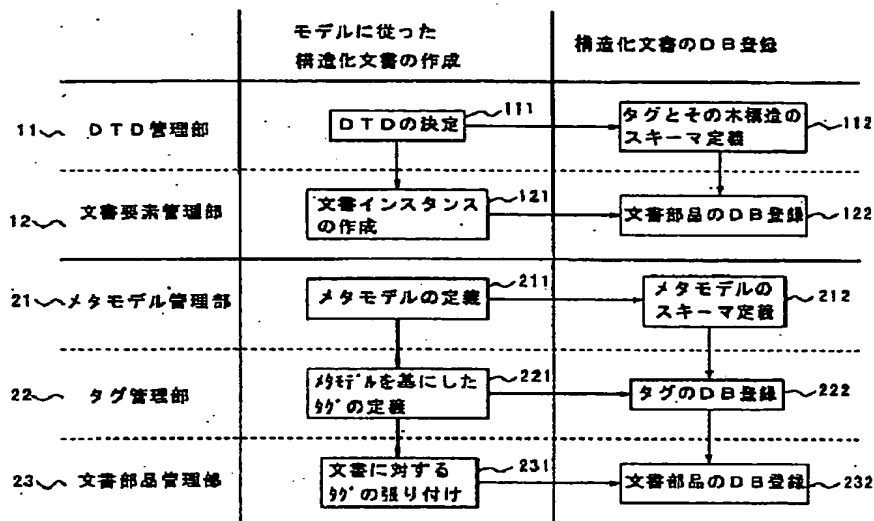
【図4】



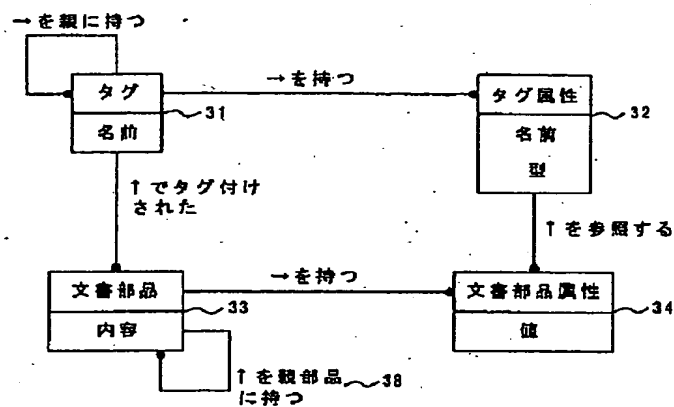
【図14】

文書部品	タグ	同義語
注文フォーム	<el id="1">	
商品	<el id="2">	
客	<el id="3">	
注文	<el id="4">	
商品名	<at id="1">	情報
価格	<at id="2">	情報
説明	<at id="3">	情報
個数	<at id="4">	
購入のfrequencyにfrequencyを入れ、注文する個数を入力する。	<at id="1">	選択表7の*1を押す

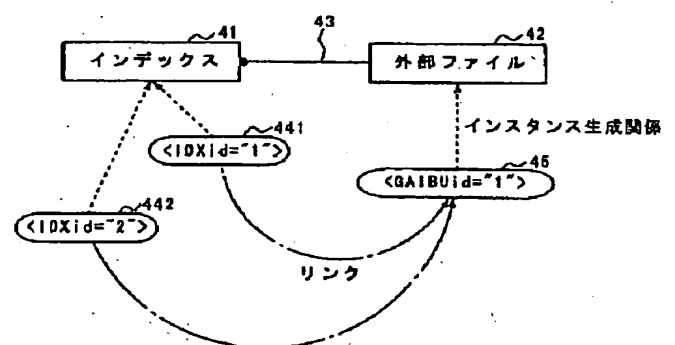
【図 3】



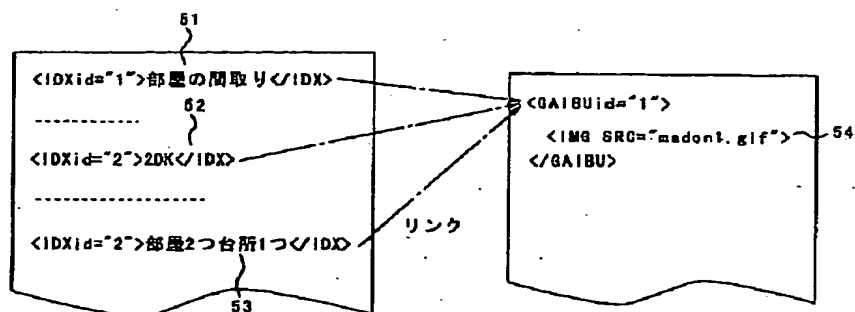
【图 5】



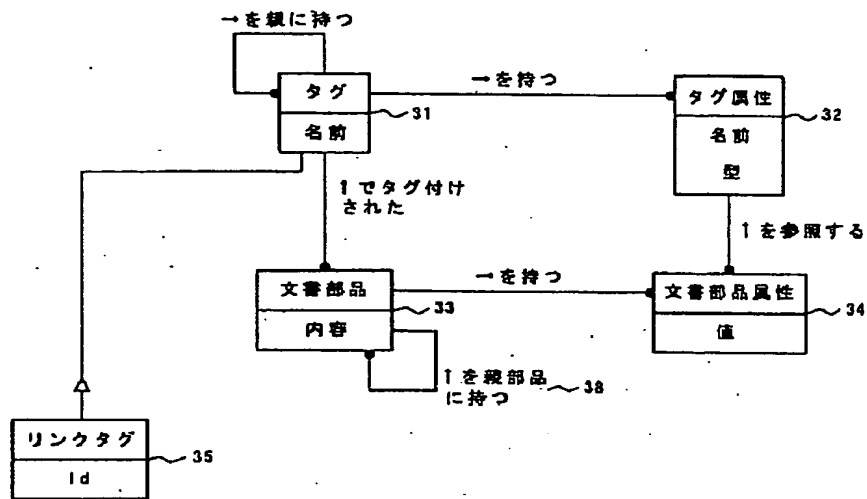
【图 8】



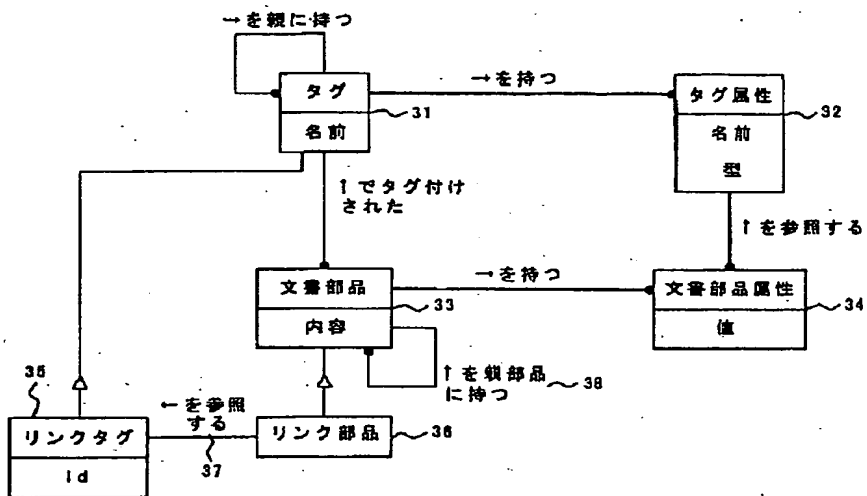
【图 9】



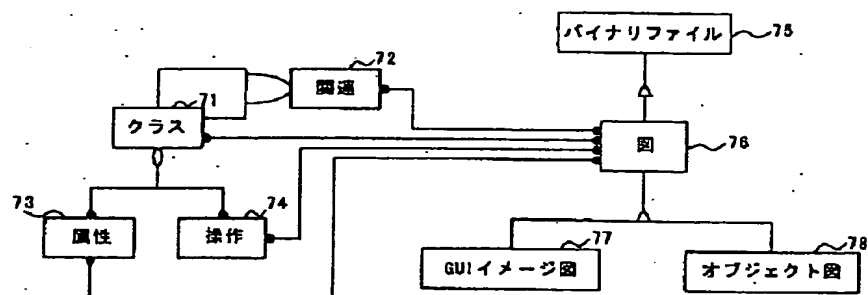
【図6】



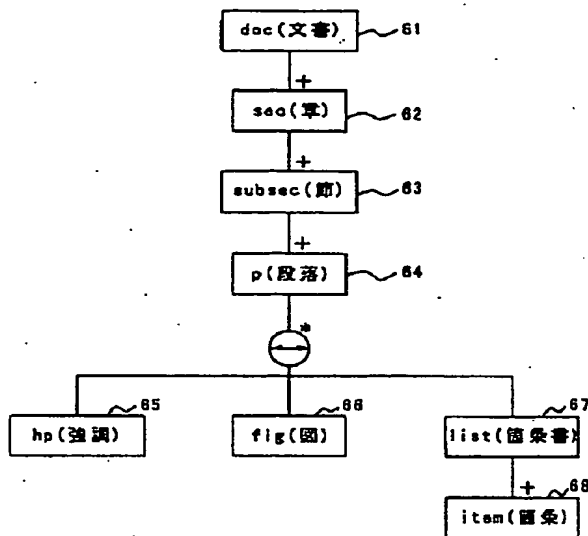
【図7】



【図12】



【図10】



【図13】

```

<doc>
  エレクトリックマスの注文処理システム
  .....
  <sec>
    基本仕様書
    .....
    <subsec>
      システムの仕様
      <p>
        <cid="1">注文フォーム</cid>にはあらかじめ扱われている
        <cid="2">商品</cid>の<cid="1"><cid="2"><cid="3">
        情報</cid></cid></cid>が商品一覧に表示されている。
        <cid="2">商品</cid>の<cid="1"><cid="2"><cid="3">
        情報</cid></cid></cid>には<cid="1">商品名</cid>、<cid="2">
        価格</cid>、そしてその<cid="2">商品</cid>の<cid="3">
        説明</cid>などがある。<cid="3">客</cid>はこれらの
        <cid="2">商品</cid>の<cid="1"><cid="2"><cid="3">
        情報</cid></cid></cid>を閲覧し、購入したい商品があったら
        <cid="1">購入ボタン</cid>にマウスをクリックし、<cid="4">注文
        </cid>する<cid="4">個数</cid>を入力する</cid>。<cid="2">
        商品</cid>の選択が終わったら、<cid="3">客</cid>は<cid="1">
        選択終了のボタンを押す</cid>
      </p>
      <p>.....</p>
      <p>.....</p>
      <p>.....</p>
    </subsec>
  </sec>
  .....
</doc>
  
```

注:「.....」は文章の省略を表す

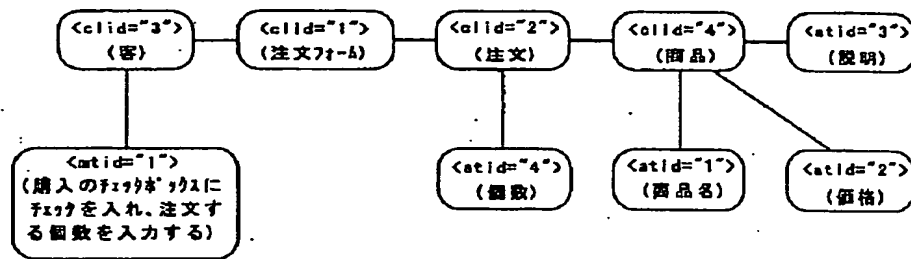
【図11】

```

<doc>
  エレクトリックマスの注文処理システム
  .....
  <sec>
    基本仕様書
    .....
    <subsec>
      システムの仕様
      <p>
        注文フォームにはあらかじめ扱われている商品の
        情報が商品一覧に表示されている。
      </p>
      <p>
        客から選択終了のボタンが押されたら、システム
        は注文された商品名、値段、注文個数、商品、と
        個数から算出される金額の小計、全注文の合計
        金額を注文フォームの注文一覧に表示する。
      </p>
      <p>
        注文一覧に記された通りに全てよければ、
        客は注文ボタンを押して注文をする。
      </p>
    </subsec>
  </sec>
  .....
</doc>
  
```

注:「.....」は文章の省略を表す

【図15】




---

フロントページの続き

(51) Int. Cl. <sup>6</sup>

識別記号

F I

G 0 6 F 15/40

3 8 0 E

15/419

3 2 0